Vitamin C and the common cold in children: severe flaws in the meta-analyses by Vorilhon et al. (2019)

Harri Hemilä[1] and Elizabeth Chalker[2]

1. Department of Public Health, University of Helsinki, Helsinki, Finland.
https://orcid.org/0000-0002-4710-307X
https://www.mv.helsinki.fi/home/hemila
harri.hemila@helsinki.fi

2. University of Sydney, Sydney, Australia
https://orcid.org/0000-0003-1593-3770

version 2021-8-18

This document describes the errors that led to the retraction of:

Vorilhon P, Arpajou B, Vaillant Roussel H, Merlin É, Pereira B, Cabaillot A.
Efficacy of vitamin C for the prevention and treatment of upper respiratory tract infection.
A meta-analysis in children.
Eur J Clin Pharmacol. 2019 Mar;75(3):303-311.
https://doi.org/10.1007/s00228-018-2601-7
https://pubmed.ncbi.nlm.nih.gov/30465062
https://www.researchgate.net/publication/329107897

THE RETRACTION NOTE

Vorilhon P, Arpajou B, Roussel HV, Merlin É, Pereira B, Cabaillot A.
Retraction Note: Efficacy of vitamin C for the prevention and treatment of upper respiratory tract infection. A meta-analysis in children.
Eur J Clin Pharmacol. 2021 Jun;77(6):941.
https://doi.org/10.1007/s00228-021-03150-9
https://pubmed.ncbi.nlm.nih.gov/33988724

SUMMARY on pages 2 to 6 describes major errors in the Vorilhon (2019) meta-analysis.

MAIN DOCUMENT on pages 7 to 34 describes details of the errors.

SUMMARY

In 2019, Philippe Vorilhon et al. published a meta-analysis on vitamin C for the prevention and treatment of upper respiratory tract infections in children [1]. In a letter to the editor we pointed out several errors in the meta-analysis [2]. Vorilhon et al. responded to our criticism [3], but we did not feel their responses were satisfactory (section 3 in the **Main Document**).

Since the publication of Vorilhon's response we delved into Vorilhon's meta-analyses in more detail. Specifics of the data extraction were not clear in Vorilhon's paper [1] and some of the trial data shown in their forest plots are questionable. We therefore reconstructed the tables from Vorilhon's meta-analysis (sections 1 and 2 of the Main Document).

The top section of the **Figure** below shows the reconstruction of Vorilhon's meta-analysis on ***common cold duration***, and the bottom section shows the meta-analysis of the same trials with the correct data. Several errors in Vorilhon's meta-analysis can be seen when comparing the two sections of the figure.

Firstly, the Cohen (2004) trial [4] is not included in the corrected meta-analysis, since it was not a vitamin C trial [2] (Main Document p 27). Secondly, in the Vorilhon analysis, there were errors in the reported numbers of observations in every trial, compare the upper and lower sections. Thirdly, in the Ritzel (1961) and Miller (1977) trials [5,6], Vorilhon used SD estimates that were inconsistent with the data published in the trial reports (Main Document pp 13,16). In this summary, we briefly describe a few of the major errors (details in the Main Document).

For older children, Coulehan (1974) reported that the number of episodes of respiratory illness was 16 in the vitamin C and 17 in the placebo groups, and the mean duration of the episodes was 4.44 days in the vitamin C and 6.29 days in the placebo groups [7]. However, instead of using the number of respiratory illness episodes as the units of observation (ie 16 and 17), Vorilhon used the number of participating children. This has the effect that Vorilhon assumed that every child suffered one episode of respiratory illness, whereas only about 13% of the children were actually infected (Main Document p 15). As a result, Vorilhon's confidence interval is much too narrow, corresponding to a P-value that is 40-times too low. Vorilhon also counted participants and not episodes in the Coulehan trial with younger children [7] and in the Ritzel trial [5] (Main Document pp 13-14).

In the Ludvigsson trial [8], the confidence interval of Vorilhon's analyses indicates that there were 11 children in the placebo group, whereas the correct number was 311 (Main Document pp 11-12). It seems that the number "3" was accidentally dropped from the beginning of the number of participants

in Vorilhon's placebo group, which would explain the very wide confidence interval. In the Miller study [6], the correct number of episodes was 220 in the vitamin C group and 211 in the placebo group, but the confidence interval indicates that only 6 observations were included in Vorilhon's calculations (Main Document p 16).

The errors described above and in the Main Document are not inconsequential. In the abstract, Vorilhon wrote that "*Vitamin C administration was found to decrease the duration of URTI by 1.6 days (standardized mean differences = -0.30 [-0.53; -0.08], p = 0.009)*" [1]. This is the pooled result shown on the top section of the Figure below. However, correct extraction of data, imputation of SD values consistent with published information, and excluding the Cohen trial that was not a vitamin C trial would lead to a revision of that statement to: "*Vitamin C administration was not found to decrease the duration of URTI (standardized mean differences = -0.15 [-0.31; +0.01], p = 0.072).*" This is the pooled result shown on the bottom section of the Figure below. Thus, correction of all the errors leads to a change from a statistically highly significant benefit to no significant difference between the treatment groups. This error is present in the abstract [1], and therefore particularly concerning.

Vitamin C may have an impact on the duration and severity of colds in children [9], but because of the severe errors in the meta-analysis, no conclusions can be drawn on this matter from Vorilhon's study [1]. Nevertheless, Vorilhon's paper is not the only one on vitamin C and respiratory infections that has been shown to be flawed. Unfortunately, many other reviews and other texts on vitamin C and respiratory infections have been shown to contain errors [10-22].

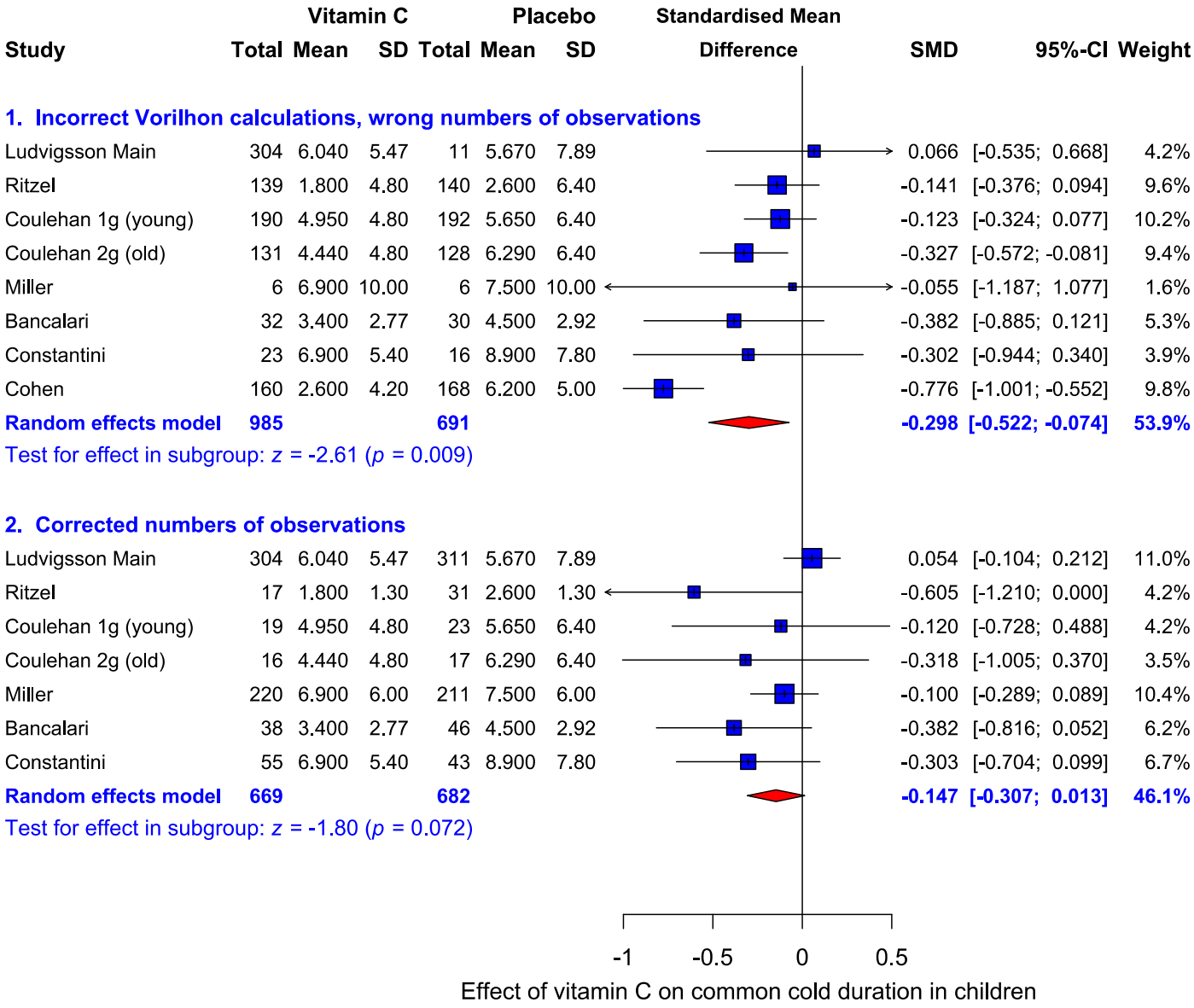


Figure. Comparison of Vorilhon's meta-analysis on the effect of vitamin C on the duration of colds with the meta-analysis with correct data. The upper panel shows the reconstructed meta-analysis on vitamin C and common cold incidence in children by Vorilhon [1]. The lower panel shows the corrected meta-analysis with data extracted from the original trial reports. See details in the Main Document. The Cohen trial [4] is not included in the corrected forest plot since participants were administered vitamin C together with echinacea and propolis. There is evidence that echinacea and propolis have an effect on the common cold [2] (Main Document p 27) and therefore the Cohen trial does not measure the effect of vitamin C alone.

References

1. Vorilhon P, Arpajou B, Vaillant Roussel H, Merlin É, Pereira B, Cabaillot A (2019) Efficacy of vitamin C for the prevention and treatment of upper respiratory tract infection: a meta-analysis in children. Eur J Clin Pharmacol 75:303-11. https://doi.org/10.1007/s00228-018-2601-7

2. Hemilä H, Chalker E (2019) Meta-analysis on vitamin C and the common cold in children may be misleading. Eur J Clin Pharmacol 75:1747-8. https://doi.org/10.1007/s00228-019-02733-x https://helda.helsinki.fi/handle/10138/318103

3. Vorilhon P, Arpajou B, Vaillant Roussel H, Merlin É, Pereira B, Cabaillot A (2019) Authors' response to letter to the editor: "Meta-analysis on vitamin C and the common cold in children may be misleading". Eur J Clin Pharmacol 75:1749-50. https://doi.org/10.1007/s00228-019-02734-w

4. Cohen HA, Varsano I, Kahan E, Sarrell EM, Uziel Y (2004) Effectiveness of an herbal preparation containing echinacea, propolis, and vitamin C in preventing respiratory tract infections in children: a randomized, double-blind, placebo-controlled, multicenter study. Arch Pediatr Adolesc Med 158:217-21. https://doi.org/10.1001/archpedi.158.3.217

5. Ritzel G (1961) Critical evaluation of vitamin C as a prophylactic and therapeutic agent in colds [in German]. Helv Med Acta 28:63-8. https://pubmed.ncbi.nlm.nih.gov/13741912 Translation: https://www.mv.helsinki.fi/home/hemila/T3.pdf

6. Miller JZ, Nance WE, Norton JA, Wolen RL, Griffith RS, Rose RJ (1977) Therapeutic effect of vitamin C: a co-twin control study. JAMA 237:248-51. https://doi.org/10.1001/jama.1977.03270300052006

7. Coulehan JL, Reisinger KS, Rogers KD, Bradley DW (1974). Vitamin C prophylaxis in a boarding school. N Engl J Med. 290:6-10. https://doi.org/10.1056/nejm197401032900102 .

8. Ludvigsson J, Hansson LO, Tibbling G (1977) Vitamin C as a preventive medicine against common colds in children. Scand J Infect Dis 9:91-8. https://doi.org/10.3109/inf.1977.9.issue-2.07 https://www.mv.helsinki.fi/home/hemila/CC/Ludvigsson_1977_ch.pdf

9. Hemilä H, Chalker E (2013) Vitamin C for preventing and treating the common cold. Cochrane Database Syst Rev CD000980. https://doi.org/10.1002/14651858.CD000980.pub4 https://helda.helsinki.fi/handle/10138/225864 https://www.mv.helsinki.fi/home/hemila/CC/2013_Coch_Colds_CD000980.pdf https://www.mv.helsinki.fi/home/hemila/CC

10. Hemilä H, Herman ZS (1995) Vitamin C and the common cold: a retrospective analysis of Chalmers' review. J Am Coll Nutr 14:116-23. https://doi.org/10.1080/07315724.1995.10718483 https://helda.helsinki.fi/handle/10138/42358 https://www.mv.helsinki.fi/home/hemila/H/HH_1995.pdf

11. Hemilä H (1996) Vitamin C supplementation and common cold symptoms: problems with inaccurate reviews. Nutrition 12:804-9. https://doi.org/10.1016/s0899-9007(96)00223-7 https://helda.helsinki.fi/handle/10138/225877 https://www.mv.helsinki.fi/home/hemila/H/HH_1996_NUT.pdf

12. Hemilä H (1996) Vitamin C, the placebo effect, and the common cold: a case study of how preconceptions influence the analysis of results. J Clin Epidemiol. 49:1079-84.
https://doi.org/10.1016/0895-4356(96)00189-8
https://helda.helsinki.fi/handle/10250/8082
https://www.mv.helsinki.fi/home/hemila/H/HH_1996_JCE.pdf

12b. Discussion (1996) J Clin Epidemiol. 49:1085-87.
https://doi.org/10.1016/0895-4356(96)00191-6
https://helda.helsinki.fi/handle/10138/225873
https://www.mv.helsinki.fi/home/hemila/H/1996_L_JCE.pdf
https://www.mv.helsinki.fi/home/hemila/H/HH_1996_JCE2.pdf

13. Hemilä H (2006) Do vitamins C and E affect respiratory infections? [Thesis] University of Helsinki, Helsinki, Finland, pp 36-45, 63-66.
https://helda.helsinki.fi/handle/10138/20335

14. Hemilä H (2009) Vitamin C for the common cold should not be rejected on the basis of old and erroneous articles. J Allergy Clin Immunol. 124:859.
https://doi.org/10.1016/j.jaci.2009.06.015
https://helda.helsinki.fi/handle/10138/16984

15. Hemilä H (2015) Zinc lozenges and vitamin C for the common cold are not examples of placebo effect in action. J Clin Epidemiol. 68:1524-5.
https://doi.org/10.1016/j.jclinepi.2015.05.012
https://helda.helsinki.fi/handle/10138/228079

16. Hemilä H (2016) Thomas Chalmers, vitamin C and the common cold. J R Soc Med. 109:46.
https://doi.org/10.1177/0141076815606279
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc4793763

17. Hemilä H (2017) Vitamin C in Clinical Therapeutics. Clin Ther. 39:2110-2.
https://doi.org/10.1016/j.clinthera.2017.08.005
https://helda.helsinki.fi/handle/10138/228957

18. Hemilä H, Chalker E (2020) Reanalysis of the effect of vitamin C on mortality in the CITRIS-ALI trial: important findings dismissed in the trial report. Front Med (Lausanne). 7:590853.
https://doi.org/10.3389/fmed.2020.590853
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7575729

19. Hemilä H, Chalker E (2021) Commentary: Vitamin C supplementation for prevention and treatment of pneumonia. Front Med (Lausanne). 7:595988.
https://doi.org/10.3389/fmed.2020.595988
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7854566

20. Hemilä H, Chalker E (2021) Commentary: The long history of vitamin C: from prevention of the common cold to potential aid in the treatment of COVID-19. Front Immunol. 12:659001.
https://doi.org/10.3389/fimmu.2021.659001
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc8047412

21. Hemilä H, Carr A, Chalker E (2021) Vitamin C may increase the recovery rate of outpatient cases of SARS-CoV-2 infection by 70%: reanalysis of the COVID A to Z randomized clinical trial. Front Immunol. 12:674681.
https://doi.org/10.3389/fimmu.2021.674681
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc8141621

22. Hemilä H, Chalker E (2021) Vitamin C and zinc lozenges for COVID-19? J Am Pharm Assoc (2003). doi: 10.1016/j.japh.2021.05.018.
https://doi.org/10.1016/j.japh.2021.05.018
https://www.ncbi.nlm.nih.gov/pmc/articles/pmc8163692

Vitamin C and the common cold in children:
severe flaws in the meta-analyses by Vorilhon et al. (2019)

Harri Hemilä and Elizabeth Chalker
https://orcid.org/0000-0002-4710-307X
https://orcid.org/0000-0003-1593-3770

MAIN DOCUMENT

Background:
In 2019, Philippe Vorilhon et al. published a meta-analysis on vitamin C and the common cold [1]. We had ourselves written a Cochrane review (2013) on vitamin C and the common cold [2], and therefore read Vorilhon's analysis with interest.
We found many problems in Vorilhon's analysis and described several of them in a letter-to-the-editor [3]. In a supplementary file to our letter, we described several further problems [4]. We did not find Vorilhon's response [5] to be satisfactory including some statements which were simply incorrect. Details of Vorilhon's calculations were not consistently provided in the meta-analyses. In this document we reconstruct Vorilhon's forest plots and show many errors in more detail.

This document has 3 parts:

Part 1 (pp 9-18) describes errors in Vorilhon's meta-analysis on vitamin C and the ***duration*** of the common cold, which was published as their figure 3 [1] (see a copy on page 10). We show that all 8 trials included in Vorilhon's meta-analysis on common cold duration either used erroneous data or should not have been included in the analysis. In their Abstract, Vorilhon wrote:
"**Vitamin C administration was found to decrease the duration of URTI by 1.6 days (standardized mean differences = -0.30 [-0.53; -0.08], p = 0.009).**"
In Part 1, we show that this is incorrect.
Correct extraction of data, imputation of SD values consistent with published information, and excluding a trial that was not a "vitamin C trial" leads to a revision of that statement to:
"**Vitamin C administration was not found to decrease the duration of URTI (standardized mean differences = -0.15 [-0.31; +0.01], p = 0.072).**"
Thus, correction of errors leads to a change from a statistically highly significant benefit to no significant difference. This error exists in the abstract of Vorilhon's paper and therefore it is particularly important.

Part 2 (pp 19-26) describes errors in Vorilhon's meta-analysis on vitamin C and the ***incidence*** of the common cold, which was published as their figure 2 [1] (see a copy on page 19).
In Part 2, we show that 5 out of 7 trials included in Vorilhon's meta-analysis on common cold incidence used erroneous data or were not studies of vitamin C.

Part 3 (pp 27-34) describes our concerns with Vorilhon's responses [5] to our critique [3,4].

Links to the references in the MAIN DOCUMENT are located on the pages where they are first cited, except that references to the first page of the MAIN DOCUMENT are listed on this page.

[1] Vorilhon P, Arpajou B, Vaillant Roussel H, Merlin É, Pereira B, Cabaillot A.
Efficacy of vitamin C for the prevention and treatment of upper respiratory tract infection.
A meta-analysis in children.
Eur J Clin Pharmacol. 2019 Mar;75(3):303-311.
https://doi.org/10.1007/s00228-018-2601-7
https://pubmed.ncbi.nlm.nih.gov/30465062
https://www.researchgate.net/publication/329107897

[2] Hemilä H, Chalker E.
Vitamin C for preventing and treating the common cold.
Cochrane Database Syst Rev. 2013 Jan 31;(1):CD000980.
https://doi.org/10.1002/14651858.CD000980.pub4
https://pubmed.ncbi.nlm.nih.gov/23440782
https://helda.helsinki.fi/handle/10138/225864
https://www.mv.helsinki.fi/home/hemila/CC/2013_Coch_Colds_CD000980.pdf
https://www.mv.helsinki.fi/home/hemila/CC

[3] Hemilä H, Chalker E.
Meta-analysis on vitamin C and the common cold in children may be misleading.
Eur J Clin Pharmacol. 2019 Dec;75(12):1747-1748
https://doi.org/10.1007/s00228-019-02733-x
https://pubmed.ncbi.nlm.nih.gov/31377890
https://www.mv.helsinki.fi/home/hemila/H/Vorilhon_2019_Problems.pdf

[4] Hemilä H, Chalker E.
Supplement to reference 3
https://static-content.springer.com/esm/228_2019_2733_MOESM1_ESM.pdf
https://www.mv.helsinki.fi/home/hemila/H/Vorilhon_2019_Problems_Supplement.pdf

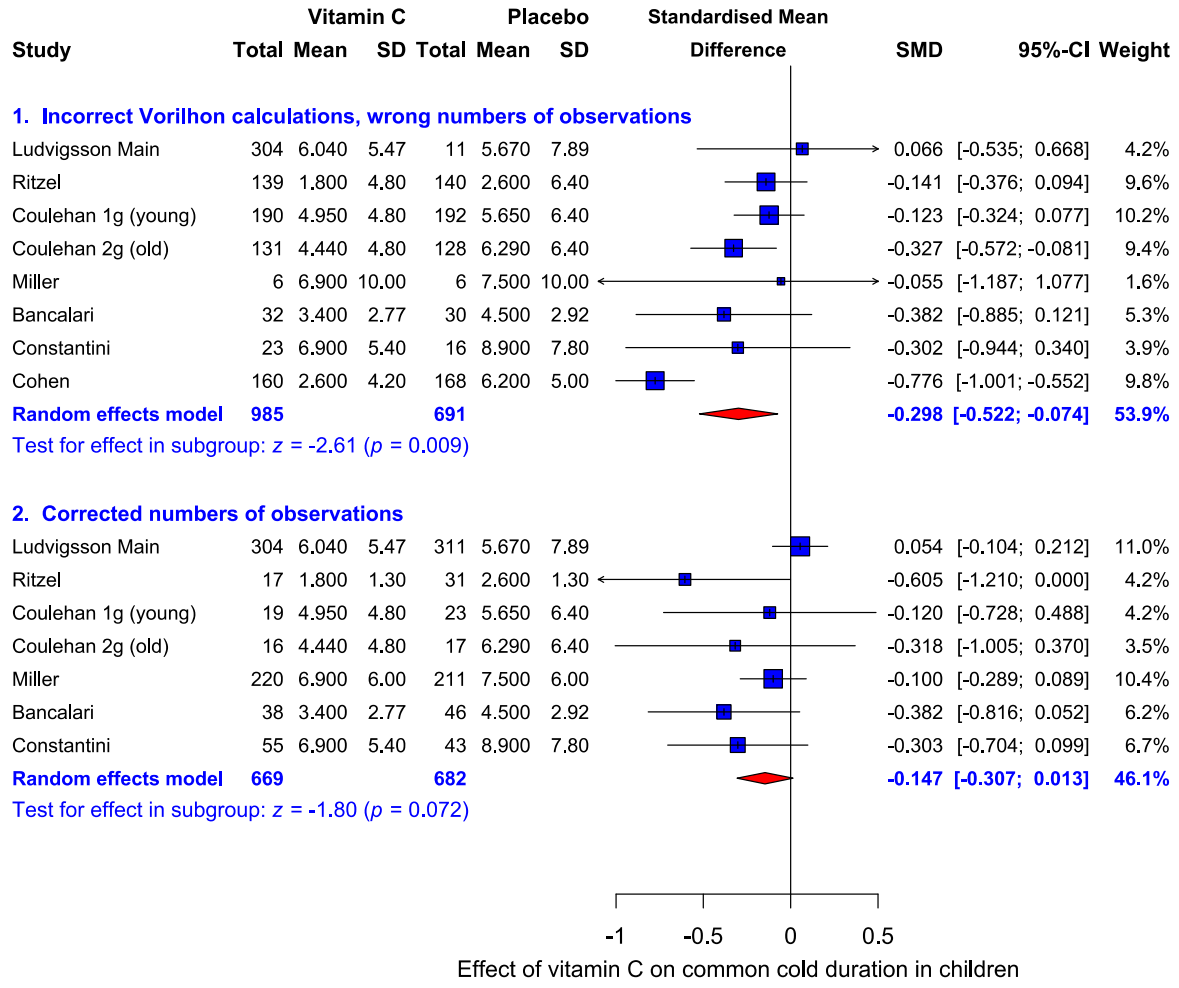
[5] Vorilhon P, Arpajou B, Vaillant Roussel H, Merlin É, Pereira B, Cabaillot A.
Authors' response to letter to the editor: "Meta-analysis on vitamin C and the common cold in children may be misleading".
Eur J Clin Pharmacol. 2019 Dec;75(12):1749-1750.
https://doi.org/10.1007/s00228-019-02734-w
https://pubmed.ncbi.nlm.nih.gov/31463576
https://www.researchgate.net/publication/335454331

Part 1: Erroneous data in Vorilhon's meta-analysis on *common cold duration*

Figure 1 below shows our reconstruction of Vorilhon's meta-analysis on common cold ***duration***. A copy of Vorilhon's figure 3 is shown on the following page as our **Figure 2**.

The bottom part of Figure 1 shows the correct analysis of the trials that were included in the Vorilhon meta-analysis. Firstly, the Cohen (2004) trial is not included, since it was not a "vitamin C trial", see [3, 4 p. 2-3] and page 27 of this document. Secondly, in the Vorilhon analysis, the number of cold episodes was incorrect for every trial. Lastly, in the Ritzel (1961) and Miller (1977) trials, Vorilhon used SD estimates that were unambiguously inconsistent with the published data. These errors are described in more detail in the following pages from 10 to 18. Note also that the correction of errors causes considerable changes in the relative weights of the trials. In Vorilhon's analysis, the Ritzel trial has 6.0 times the weight of the Miller trial (9.6/1.6). Once corrected, the Miller trial has 2.5 times the weight of the Ritzel trial (10.4/4.2). Thus, there is a 15 fold difference between the weights of these two trials in the Vorilhon meta-analysis and in the corrected meta-analysis. In Figure 1 and in the discussion below, the trials are ordered by the severity of errors in Vorilhon's meta-analysis.

Figure 1: Reconstruction of Vorilhon's figure 3 on the top and corrected numbers of episodes at the bottom.



Vorilhon's figure 3 [1]

Figure 2 below shows Vorilhon's meta-analysis on common cold duration. **Figure 3** below shows our reconstruction in which the trials are in the same order as in Vorilhon's meta-analysis. The pooled 95% CIs are very similar in both figures. In addition, both give P = 0.009 for a test of the overall difference. There are minor differences in the 95% CIs of the individual trials but these are likely due to rounding.

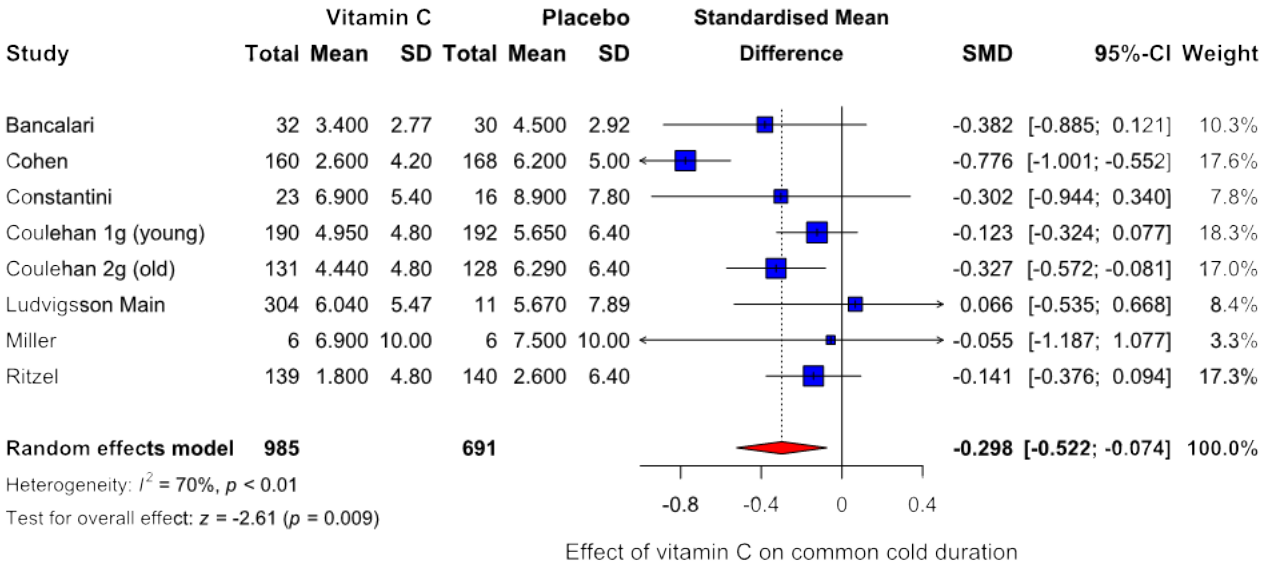
Vorilhon reported the mean duration and its SD, but the number of observations was not shown. In our reconstruction we searched for the number of observations that led to the 95% CI published by Vorilhon. Vorilhon's errors in the number of observations, in the imputed SD values, and in the calculated effect estimates are indicated on the following pages by yellow with explanations.

Figure 2: A copy of Vorilhon's figure 3. The red underlining indicates the pooled effect.

Study name	Statistics for each study					Mean ± SD	
	Std diff in means	Standard error	Lower limit	Upper limit	p-value	Treated	Control
Bancarali et al. (1984)	-0.387	0.256	-0.890	0.116	0.132	3.40 ± 2.77	4.50 ± 2.92
Cohen et al. (2004)	-0.778	0.115	-1.003	-0.553	0.000	2.60 ± 4.20	6.20 ± 5.00
Constantini et al. (2010)	-0.302	0.323	-0.936	0.331	0.349	6.90 ± 5.40	8.90 ± 7.80
Coulehan et al. (1974, 1g)	-0.124	0.102	-0.324	0.077	0.227	4.95 ± 4.80	5.65 ± 6.40
Coulehan et al. (1974, 2g)	-0.328	0.125	-0.573	-0.082	0.009	4.44 ± 4.80	6.29 ± 6.40
Ludvigsson et al. (1977)	0.067	0.307	-0.535	0.668	0.828	6.04 ± 5.47	5.67 ± 7.89
Miller et al. (1977)	-0.060	0.606	-1.247	1.127	0.921	6.90 ± 10.00	7.50 ± 10.00
Ritzel et al. (1961)	-0.141	0.120	-0.376	0.094	0.238	1.80 ± 4.80	2.60 ± 6.40
	-0.300	0.115	-0.525	-0.076	0.009		

Fig. 3 Forest plot: duration of upper respiratory tract infection

Figure 3: Reconstructed Vorilhon meta-analysis on duration:



Ludvigsson (1977) [6] "*Main study*" data in table V is copied to our **Figure 4** on the next page.

For the "cold symptoms from the nose" Ludvigsson reported $t = 0.67$, which is consistent with our analysis below.

Correct analysis of the Ludvigsson (1977) Main study trial for the duration of colds

Ludvigsson Main	N	Mean duration	SD	SMD (95% CI)
Vitamin C	304	6.04	5.47	0.054 (-0.104,+0.213) Z = 0.67
Placebo	311	5.67	7.89	
Unit of analysis	Participant			

Vorilhon's incorrect analysis of the Ludvigsson (1977) Main study for the duration of colds

Ludvigsson Main	N	Mean duration	SD	SMD (95% CI)
Vitamin C	304	6.04	5.47	0.066 (-0.535,+0.668) Z = 0.22
Placebo	**11** [a)]	5.67	7.89	
Unit of analysis	Participant			

[a)] The difference in Vorilhon's estimate (SMD = 0.067) and our estimate (SMD = 0.054) indicated that there needs to be substantial difference in the size of the two groups to explain the change in SMD. "304" is the number of participants in the vitamin C group, and "311" is the correct number of participants in Ludvigsson's placebo group (Figure 4). Thus, it seems likely that Vorilhon accidentally dropped the number "3" from the beginning of the number of participants in the placebo group, which would explain the "11" and the very wide resulting confidence interval. The Z-value for Vorilhon's comparison is inconsistent with Ludvigsson's report.

Using the above numbers of participants (304 and 11) gives **exactly** the same 95% CI that was reported by Vorilhon (Figure 2). **Once corrected, the much narrower confidence interval means that the trial has a much greater weight in the pooled analysis, see Figure 1.**

Furthermore, in our table above we show the number of participants, since thereby we can illustrate most clearly the error in Vorilhon's analysis. However, the more appropriate statistical analysis would be by the number of common cold episodes. There were about 2 episodes per participant (next page).

Finally, the title of Vorilhon's [1] paper included the term "upper respiratory tract infection" [URTI] and that term was repeated in the text. However, for the Ludvigsson (1977) trial, Vorilhon extracted duration data only for "Cold symptoms from the nose", even though outcome data for URTI was also available, see **Figure 4**. No justification was given for selecting only nose colds.

[6] https://pubmed.ncbi.nlm.nih.gov/897573
https://doi.org/10.3109/inf.1977.9.issue-2.07
https://www.mv.helsinki.fi/home/hemila/CC/Ludvigsson_1977_bm.pdf
https://www.mv.helsinki.fi/home/hemila/CC/Ludvigsson_1977_ch.pdf

Figure 4: A copy of Ludvigsson's table V

Table V. *Occurrence of certain cold variables in control group and vitamin C group*

	Totally free from symptoms			Incidence (no. of cases/person)		Duration (no. of days/period)	
	N	%	*t*	M±S.D.	*t*	M±S.D.	*t*
Pilot study (30 mg, *N*=78; 1 000 mg, *N*=80)							
Cold symptoms from the nose							
30 mg	17	22	0.87	1.36±1.21	−1.59	7.61±8.07	1.82
1 000 mg	13	16		1.63±1.15		5.39±4.88	
Upper respiratory tract infection							
30 mg	34	44	0.68	0.71±0.72	−0.72	14.53±9.75	3.05**
1 000 mg	31	39		0.78±0.75		8.90±5.96	
Main study (10 mg, *N*=311; 1 000 mg, *N*=304)							
Cold symptoms from the nose							
10 mg	53	17	0.37	2.00±1.80	−1.41	5.67±7.89	−0.67
1 000 mg	49	16		2.16±1.63		6.04±5.47	
Upper respiratory tract infection							
10 mg	71	23	−0.35	1.28±1.03	−1.38	10.14±11.60	0.56
1 000 mg	74	24		1.39±1.11		9.54±8.65	

Number of participants

Vorilhon's duration analysis was based on "Cold symptoms from the nose"

Vorilhon uses the term "upper respiratory tract infection" [URTI] even in the title of their paper [1], but they did not extract duration of URTI although it was available as a separate outcome.

Ritzel (1961) [7] reported the duration of colds in the text section:

1.8 days mean cold duration in the vitamin C group
2.6 days mean cold duration in the placebo group

The 1.8 days mean common cold duration is based on 17 participants who became sick (see p 20). The 2.6 days mean common cold duration in based on 31 participants who became sick (see p 20). Since the duration of the trial was about 1 week, the number of sick participants is equal to the number of common cold episodes.

This logic leads to the following data:

Correct analysis of the Ritzel (1961) trial for the duration of colds

Ritzel	N	Mean duration	SD	SMD (95% CI)
Vitamin C	17	1.8	1.3 [a)]	-0.605 (-1.21, 0.00) P = 0.050
Placebo	31	2.6	1.3 [a)]	
Unit of analysis	Common cold episode			

[a)] The above SD estimate of 1.3 days is based on imputing a SD value that is conservatively consistent with P < 0.05, which was reported by Ritzel (1961) for the comparison of the vitamin C and placebo groups.

Vorilhon's incorrect analysis of the Ritzel (1961) trial for the duration of colds

Ritzel	N	Mean duration	SD	SMD (95% CI)
Vitamin C	139	1.8	4.8 [b)]	-0.141 (-0.376, +0.094) P = 0.238
Placebo	140	2.6	6.4 [b)]	
Unit of analysis	Participant			

[b)] These SD estimates were used by Vorilhon, but no description was given as to how they were imputed [1,5]. Previously, we pointed out that Vorilhon's SDs of 4.8 and 6.4 must be incorrect since the resulting P = 0.238 is inconsistent with the P-value published by Ritzel [4 p. 13-14].

Using the number of participants (139 and 140) gives **exactly** the same 95% CI reported by Vorilhon (Figure 2). Thus, Vorilhon did not use common cold episodes as the units of analysis, but participants. However, few participants caught the common cold (12% [17/139] in vitamin C and 22% [31/140] in placebo groups) and therefore using participants rather than common cold episodes leads to a completely incorrect analysis; it means **counting non-sick people as sick**.

Furthermore, because of the arbitrary SD imputations, **the pooled SMD estimate calculated by Vorilhon is one quarter of the correct SMD estimate**. Note that the 95% CI is much wider using cold episodes rather than participants and with the SD that is consistent with Ritzel's report of P < 0.05. **The much wider CI results in the trial having far less weight when pooled (Figure 1)**.

[7] https://www.mv.helsinki.fi/home/hemila/T3.pdf (Translation)
https://pubmed.ncbi.nlm.nih.gov/13741912
https://doi.org/10.1001/jama.1976.03260370018017
https://www.mv.helsinki.fi/home/hemila/CC/Ritzel_1961_ch.pdf
https://www.mv.helsinki.fi/home/hemila/CC/Ritzel_1961_bm.pdf
https://www.mv.helsinki.fi/home/hemila/CC/Ritzel_1976.pdf

Coulehan (1974) Young [8] reported the "*Average Duration of [Respiratory illness] Episodes*" and the "*No. of [Respiratory illness] Episodes*" in their *table 2*:

Young children (1g, Lower grades):

4.95 days mean cold duration in the vitamin C group with 19 cold episodes
5.65 days mean cold duration in the placebo group with 23 cold episodes

This leads to the following data:

Correct analysis of the Coulehan (1974) trial for the duration of colds

Coulehan young	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	19	4.95	4.95 [a)]	-0.120 (-0.73, +0.49)
Placebo	23	5.65	5.65 [a)]	
Unit of analysis	Common cold episode			

[a)] This imputation follows our approach in our Cochrane review (2013) [2]: "*In the other trials with missing SD, we estimated SD as identical with the mean of the treatment group. This is based on our analysis that for trials reporting the SD, the ratio of SD to mean is on average 0.7 so that our ratio of 1.0 used in the SD imputation is somewhat conservative. The consequence of this is that we are putting slightly reduced weight in our estimates of effect on these trials with missing SD values, compared to the average.*"

Vorilhon's incorrect analysis of the Coulehan (1974) trial for the duration of colds

Coulehan young	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	190	4.95	4.8 [b)]	-0.123 (-0.324, +0.077)
Placebo	192	5.65	6.4 [b)]	
Unit of analysis	Participant			

[b)] These are the same SD values Vorilhon imputed for Ritzel. They are unambiguously incorrect for the Ritzel trial, but not far from those we imputed ourselves above. Therefore we are not sure whether they are correct or not, given that Vorilhon did not provide any explanation of how they were imputed.

Using the number of participants (190 and 192) gives **exactly** the same 95% CI that was reported by Vorilhon (Figure 2). Thus, Vorilhon did not use common cold episodes as the unit of analysis, but the number of participants. However, few participants caught the common cold (10% [19/190] in vitamin C and 12% [23/192] in placebo groups) and therefore the use of participants again leads to an incorrect analysis in which **people who were not sick were counted as sick**.
Note that the 95% CI is much wider using cold episodes rather than participants. **The much wider CI results in the trial having far less weight when pooled (Figure 1).**

[8] https://doi.org/10.1056/NEJM197401032900102
https://pubmed.ncbi.nlm.nih.gov/4586102

Coulehan (1974) Old [8] reported the "*Average Duration of [Respiratory illness] Episodes*" and the "*No. of [Respiratory illness] Episodes*" in their *table 2*:

Old children (2g, Upper grades):

4.44 days mean cold duration in the vitamin C group with 16 cold episodes
6.29 days mean cold duration in the placebo group with 17 cold episodes

This leads to the following data:

Correct analysis of the Coulehan (1974) trial for the duration of colds

Coulehan old	N	Mean duration	SD	SMD (95% CI)
Vitamin C	16	4.44	4.44 [a)]	-0.318 (-1.01,+0.37) P = 0.36
Placebo	17	6.29	6.29 [a)]	
Unit of analysis	Common cold episode			

[a)] See Coulehan young children on the previous page.

Vorilhon's incorrect analysis of the Coulehan (1974) trial for the duration of colds

Coulehan old	N	Mean duration	SD	SMD (95% CI)
Vitamin C	131	4.44	4.8 [b)]	-0.327 (-0.572, -0.081) P = 0.009
Placebo	128	6.29	6.4 [b)]	
Unit of analysis	Participant			

[b)] See Coulehan young children on the previous page.

Using the number of participants gives **very close** to the same SMD and 95% CI that was reported by Vorilhon (Figure 2). Thus, it seems evident that Vorilhon did not use common cold episodes as the units of analysis, but the number of participants. However, few participants caught the common cold and therefore the use of participants leads to grossly false analysis. **It means counting nonsick people as sick.**
Note that the 95% CI is much wider using cold episodes rather than participants. **The much wider CI causes that the trial has much less weight in correct pooling (Figure 1).**
In addition, for the Coulehan (1974) older children, **Vorilhon calculated P = 0.009 (see Figure 2 on page 10 of this Supplement). However, the correct value is P = 0.36. Thus, Vorilhon's P-value is 40-times too low.**

[8] https://doi.org/10.1056/NEJM197401032900102
https://pubmed.ncbi.nlm.nih.gov/4586102

Miller (1977) studied twins [9]. The number of participants was 44 *twin pairs* as reported in their *table 1*. Miller's *table 3* reports the number of common cold episodes per group.

Total number of colds is the product of the incidence of colds and the number of participants:
Vitamin C: 5.0 × 44 = 220 common cold episodes.
Placebo: 4.8 × 44 = 211 common cold episodes.

Miller also reported that colds were 0.6 days (SE 0.6 days) shorter in the 44 twins receiving vitamin C.
For the t-test, Z = mean/SE = -0.6/0.6 = -1.0.
This gives P (2-tail, Z = -1.0) = 0.32.
Thus, the imputed SD values must be consistent with a Z value close to 1.0.

These data lead to following analysis:

Correct analysis of the Miller (1977) trial for the duration of colds

Miller	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	220	6.9	6.0 [a)]	-0.100 (-0.289, +0.089) Z = 1.04
Placebo	211	7.5	6.0 [a)]	
Unit of analysis	Common cold episode			

[a)] The SD estimates were imputed to be consistent with Z = 1.0 in the t-test.

Vorilhon's incorrect analysis of the Miller (1977) trial for the duration of colds

Miller	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	6 [b)]	6.9	10 [c)]	-0.055 (-1.187, +1.077) Z = 0.10
Placebo	6 [b)]	7.5	10 [c)]	
Unit of analysis	Participant			

[b)] The total number of pairs was 44. The number of participants ranged from 5 to 13 in the six boy/girl groups administered 0.5, 0.75, and 1.0 g/day vitamin C. The number of boys given 0.75 g/day vitamin C was "6". It seems that Vorilhon's analysis did not include all 44 pairs of twins, but just one of the groups defined by sex and dose.
[c)] Vorilhon does not provide any explanation of how the SDs were estimated, though Miller's report does give relevant information (see above). Vorilhon's SD = 10 leads to Z = 0.10, whereas Miller reported Z = 0.6/0.6 = 1.0.

Using the 6 participants gives **very closely** the same SMD and 95% CI that was reported by Vorilhon (Figure 2). Thus, Vorilhon did not use common cold episodes as the units of analysis, but the number of participants apparently from just one of the groups. Note that the 95% CI is much narrower using all observed cold episodes rather than a subgroup of participants. **The much narrower CI causes that the trial has much greater weight in correct pooling (Figure 1).**

[9] https://doi.org/10.1001/jama.1977.03270300052006
https://pubmed.ncbi.nlm.nih.gov/318715

Bancalari (1984) [10] reported the number and duration of colds as follows:
"*In the placebo group (n = 30) 46 episodes of ARI were detected vs 38 episodes of ARI in the Vitamin C group (n = 32). Vitamin C group presented a significant decrease (p < 0.05) in duration of ARI (3.4 days ± 0.45 S.D.) as compared to placebo group (4.5 days ± 0.43).*"

The term "SD" is incorrect and actually indicates "SE".
The SD can be calculated a follows (SD = √n × SE):

Vitamin C group: SD = √38 × 0.45 = 2.773.
Placebo C group: SD = √46 × 0.43 = 2.916.

This leads to the following data:

Correct analysis of the Bancalari (1984) trial for the duration of colds

Bancalari	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	38	3.4	2.77	-0.382 (-0.816,+0.052)
Placebo	46	4.5	2.92	
Unit of analysis	Common cold episode			

Vorilhon's incorrect analysis of the Bancalari (1984) trial for the duration of colds

Bancalari	**N**	**Mean duration**	**SD**	**SMD (95% CI)**
Vitamin C	32	3.4	2.77	-0.382 (-0.885, +0.121)
Placebo	30	4.5	2.92	
Unit of analysis	Participant			

Using the number of participants gives **very close** to the same SMD and 95% CI that was reported by Vorilhon (Figure 2). Thus, Vorilhon did not use common cold episodes as the units of analysis, but the number of participants. In this case, the error does not cause as dramatic a change in the 95% CI as in the trials by Ritzel, Coulehan and Miller.

[10] https://www.mv.helsinki.fi/home/hemila/T6.pdf (Translation)
https://pubmed.ncbi.nlm.nih.gov/6398492
https://www.mv.helsinki.fi/home/hemila/CC/Bancalari_1984_ch.pdf
https://www.mv.helsinki.fi/home/hemila/CC/Bancalari_1984_bm.pdf

Constantini (2011) [11]

Constantini's results which also show the number of common cold episodes used in the analysis is published in their *table 3*.

Correct analysis of the Constantini (2011) trial for the duration of colds

Constantini	N	Mean duration	SD	SMD (95% CI)
Vitamin C	55	6.9	5.4	-0.303 (-0.704,+0.099)
Placebo	43	8.9	7.8	
Unit of analysis	Common cold episode			

Vorilhon's incorrect analysis of the Constantini (2011) trial for the duration of colds

Constantini	N	Mean duration	SD	SMD (95% CI)
Vitamin C	23	6.9	5.4	-0.302 (-0.944, +0.340)
Placebo	16	8.9	7.8	
Unit of analysis	Participant			

Using the number of participants gives **very close** to the same SMD and 95% CI that was reported by Vorilhon (Figure 2). Thus, Vorilhon did not use common cold episodes as the units of analysis, but the number of participants. In this case, the error does not cause as dramatic a change in the 95% CI as in the trials by Ritzel, Coulehan and Miller.

[11] https://doi.org/10.1007/s00431-010-1270-z
https://pubmed.ncbi.nlm.nih.gov/20689965
https://helda.helsinki.fi/handle/10138/228085

Part 2: Incorrect data in Vorilhon's meta-analysis on *common cold incidence*

Figure 5 shows a copy of Vorilhon's [1] meta-analysis on common cold ***incidence*** which was published as their figure 2.

In their figure 2, Vorilhon calculated odds ratio (OR) estimates for the effect of vitamin C on common cold incidence. In our first critique we pointed out that OR is a poor measure for outcomes that are common, and there is also no need to use OR in the analysis of randomized trials, since the risk ratio (RR) can be directly calculated from RCT data [4 p. 12].

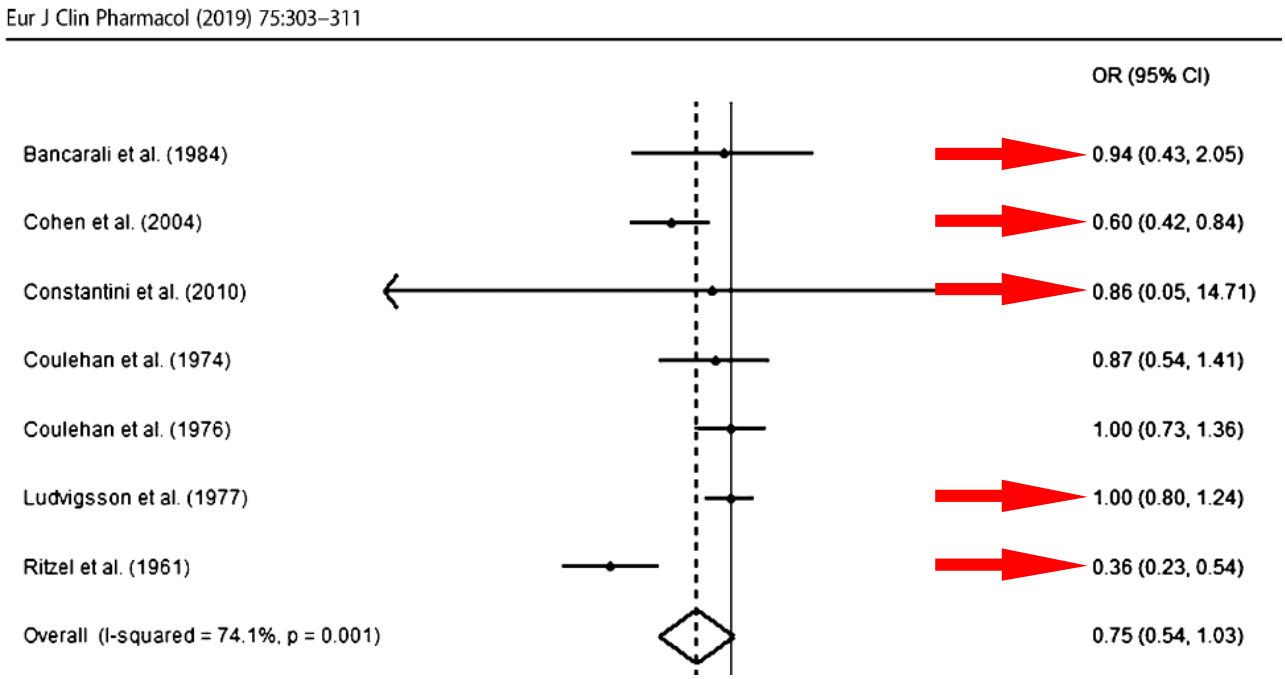
OR is calculated as the ratio of sick vs. non-sick in the vitamin C vs placebo groups:

For a 2×2 table on the right-hand side (sick/not sick vs. vitamin C/placebo)

	Total	**Sick**	**Not sick**
Vitamin C	a + b	a	b
Placebo	c + d	c	d

The calculation is as follows: OR = (a/b) / (c/d)

Figure 5: A copy of Vorilhon's figure 2 [1]. Red arrows indicate incorrect values. The Cohen trial was not a "vitamin C study", see [3, 4 p. 2-3] and page 27 of this document. Errors in the Ritzel, Bancalari, Ludvigsson, and Constantini trials are described on the following pages. The correct data and the correct OR estimates are shown on pages 20-24 of this document, together with the incorrect data that Vorilhon used in the calculation of their OR values.



Ritzel (1961) [7] reported the number of days sick (Krankheitstage) and the total number of symptoms (Einzelsymptome) in *tabelle 1*, which is shown as our **Figure 6** below.
The number of sick participants can be calculated from total days sick per group by dividing by the mean days of sickness which was reported in the text section [7; translation available]:

1.8 days mean cold duration in the vitamin C group: 31/1.8 = 17 as the closest integer
2.6 days mean cold duration in the placebo group: 80/2.6 = 31 as the closest integer

Figure 6: A copy of Ritzel's table 1

G. Ritzel: Vitamin C bei Erkältungskrankheiten

Tabelle 1
Erkältungsprophylaxe durch Vitamin C (1,0 g täglich): Häufigkeit der Krankheitstage und Einzelsymptome im Vergleich zu einem scheinbehandelten Kollektiv

	Mit Vitamin C behandelt	Placebo verabreicht	Total
Anzahl Versuchsteilnehmer	139	140	279
Anzahl Krankheitstage	31	80	111
Anzahl Einzelsymptome	42	119	161

From the total number of participants and the number of sick, the number of non-sick can be calculated and thereby the 2×2 table can be constructed on the right-hand side of the table.

Correct Ritzel (1961) results:

Ritzel	Total	Sick	Nonsick
Vitamin C	139	17	122
Placebo	140	31	109

Correct OR(Ritzel) = (17/122) / (31/109) = 0.490

However, Vorilhon published OR(Ritzel) = 0.36 (95% CI 0.23, 0.54), which is incorrect.
In our critique, we already showed that Vorilhon's OR for the Ritzel trial was incorrect [4 p. 13]. However, Vorilhon [5] did not respond to the concern we expressed. We therefore decided to determine the data set from which Vorilhon calculated the incorrect OR value and its 95% CI.

Vorilhon's OR estimate can be explained with the following data:

Ritzel	Total	Sick	Nonsick
Vitamin C	181	42	139
Placebo	259	119	140

This gives the incorrect Vorilhon's OR(Ritzel) = (42/139) / (119/140) = 0.355
The above table gives **exactly** the same 95% CI that was reported by Vorilhon (Figure 5 and p. 19). When extracting the data, it seems that Vorilhon misunderstood that **Anzahl Einzelsymptome** means the number of sick people. He seems also to have understood (mistakenly) that 139 and 140 indicate the non-sick people, whereas they are actually the totals in the trial groups.

Bancalari (1984) [10] reported the number of children with no common cold during the trial in *Tabla 3*, which is shown as our **Figure 7**.

Figure 7: A copy of Bancalari's tabla 3

Tabla 3. Efecto de la Vitamina C vs placebo en el número de infecciones respiratorias agudas durante el período estudiado

Grupo	sin episodios de IRA n	%
Vitamina C (n = 32)	11	34,3
Placebo (n = 30)	9	30,0

This data allows us to populate the 2×2 table for calculating the OR.

Correct Bancalari (1984) results:

Bancalari	Total	Sick	Not sick
Vitamin C	32	21	11
Placebo	30	21	9

Correct OR(Bancalari) = (21/11) / (21/9) = 0.818

However, Vorilhon published OR(Bancalari) = 0.94 (95% CI 0.43, 2.05), which is incorrect.

Vorilhon's OR estimate can be explained with the following data:

Bancalari	Total	Sick	Not sick
Vitamin C	53	21	32 (=21+11)
Placebo	51	21	30 (=21+9)

This gives the Vorilhon's incorrect OR(Bancalari) = (21/32) / (21/30) = 0.938

The above table gives **exactly** the same 95% CI that was reported by Vorilhon (Figure 5 and p. 19). When extracting the data, it seems that Vorilhon has misunderstood that 32 and 30 indicate the non-sick people, whereas they indicate the total number of participants in the trial groups. **The sick people are double counted**. Thus, the 21 sick participants are also counted as non-sick participants in Vorilhon's analysis.

Ludvigsson (1977) [6] published their "*Pilot study*" and "*Main study*" in Table V, a copy of which is shown in our **Figure 8** on the next page.

The published data lead to two tables for the calculation of the two separate OR values.
Pilot study:

Ludvigsson Pilot	Total	Sick	Nonsick
Vitamin C	80	49	31
Placebo	78	44	34

Correct OR(Ludvigsson Pilot) = (49/31) / (44/34) = 1.221

Main study:

Ludvigsson Main	Total	Sick	Nonsick
Vitamin C	304	230	74
Placebo	311	240	71

Correct OR(Ludvigsson Main) = (230/74) / (240/71) = 0.919

A meta-analysis should analyze the above two different trials separately. Nevertheless, if the Pilot and Main studies are combined, the resulting table is as follows:

Ludvigsson Combined	Total	Sick	Nonsick
Vitamin C	384	279	105
Placebo	389	284	105

Correct OR(Ludvigsson Combined) = (279/105) / (284/105) = 0.982

Vorilhon combined the two separate Ludvigsson trials.
However, Vorilhon published OR(Ludvigsson) = 1.00 (95% CI 0.80, 1.24), which is incorrect.
Vorilhon's confidence interval is much too narrow.

Vorilhon's OR estimate can be explained with the following data:

Ludvigsson Combined	Total	Sick	Nonsick
Vitamin C	663	279	384 (=279+105)
Placebo	673	284	389 (=284+105)

This gives Vorilhon's OR(Ludvigsson Combined) = (279/384) / (284/389) = 0.995
The above table gives **exactly** the same 95% CI that was reported by Vorilhon (Figure 5 and p. 20).

When extracting the data, it seems that Vorilhon mistakenly interpreted that 384 and 389 were non-sick, whereas they are the total number of participants in the combined vitamin C and placebo groups. The 279 and 284 were actually reported in Vorilhon's response [5]: "*The data were pooled from the two studies (279/384 vs. 284/389 ...*". The "279" and "284" are correct combined numbers for the sick. However, in Vorilhon's analysis those **sick people are double counted** also as non-sick in the same way as in the Bancalari study, see previous page. Finally, for the incidence, Vorilhon extracted incidence data for "upper respiratory infections" (see p 17), but duration data for "cold symptoms from the nose" (see p 6) without explaining the inconsistency.

Figure 8: a copy of Ludvigsson's table V

Table V. *Occurrence of certain cold variables in control group and vitamin C group*

	Totally free from symptoms			Incidence (no. of cases/person)		Duration (no. of days/period)	
	N	%	*t*	M±S.D.	*t*	M±S.D.	*t*
Pilot study (30 mg, *N*=78; 1 000 mg, *N*=80)							
Cold symptoms from the nose							
30 mg	17	22	0.87	1.36±1.21	−1.59	7.61±8.07	1.82
1 000 mg	13	16		1.63±1.15		5.39±4.88	
Upper respiratory tract infection							
30 mg	34	44	0.68	0.71±0.72	−0.72	14.53±9.75	3.05**
1 000 mg	31	39		0.78±0.75		8.90±5.96	
Main study (10 mg, *N*=311; 1 000 mg, *N*=304)							
Cold symptoms from the nose							
10 mg	53	17	0.37	2.00±1.80	−1.41	5.67±7.89	−0.67
1 000 mg	49	16		2.16±1.63		6.04±5.47	
Upper respiratory tract infection							
10 mg	71	23	−0.35	1.28±1.03	−1.38	10.14±11.60	0.56
1 000 mg	74	24		1.39±1.11		9.54±8.65	

Vorilhon duration analysis was based on "Cold symptoms from the nose"

Vorilhon incidence analysis was based on "Upper respiratory tract infection"

Constantini (2011) [11] published the number of colds in the vitamin C and placebo groups in their *table 2*, data from which is copied below.

Correct Constantini (2011) results:

Constantini	N Participants	N episodes	RR (95% CI)
Vitamin C	21	64	1.01 (0.70 – 1.46)
Placebo	18	54	

OR is a useful approximation for risk ratio (RR) when the number of cases is low [4 p. 12].
For example, Altman wrote that "*the odds ratio [OR] should not be interpreted as an approximate relative risk [RR] unless the events are rare in both groups (say, less than 20–30%)*" [12].

When the disease is so common that there are about 3 episodes per participant as in the table above, and essentially all participants suffer from the disease, OR is definitely an inappropriate method for approximating the RR.

Vorilhon published 95% CI: 0.05 – 14.71 without any description of the data on which that calculation was based (Figure 5). That means that, according to Vorilhon, the results by Constantini (2011) are consistent with vitamin C decreasing common cold incidence by 95% or increasing common cold incidence by 1371%. Vorilhon's confidence interval gives a very misleading impression of the (in)accuracy of the Constantini (2011) trial. The correct confidence interval for the RR which is shown in the table above extends from a 30% decrease in common cold incidence to a 46% increase in incidence.

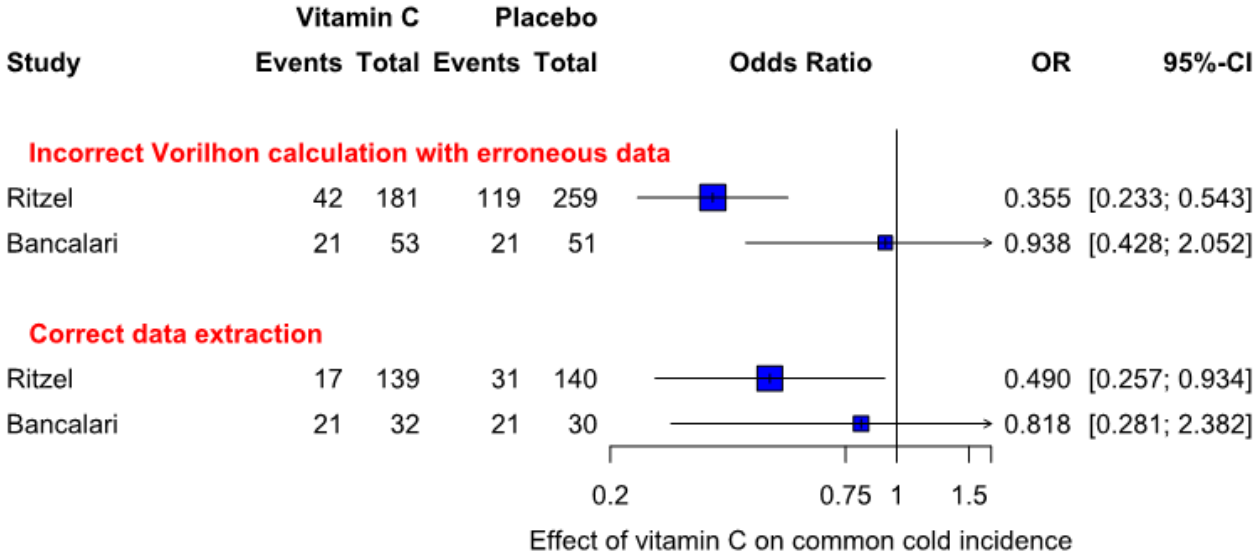
When a study cannot be included in the meta-analysis, its results could nevertheless be described in the text section with the 95% CI of the effect measure that was used in the original publication (ie RR). There is no basis for calculating an OR value for the Constantini (2011) trial and including it in the meta-analysis of Vorilhon's figure 2.

[12] https://www.bmj.com/content/317/7168/1318.1

Calculation of the 95% CI for the Ritzel (1961) and Bancalari (1984) trials

Figure 9 shows Vorilhon's calculations on the top and the correct published data and the corresponding correct OR and 95% on the bottom. Compare against Vorilhon's figure 2, a copy of which is shown as our Figure 5 (p. 13).

Figure 9: Comparison of correct data extraction from the Ritzel and Bancalari trials with the incorrect data extraction by Vorilhon

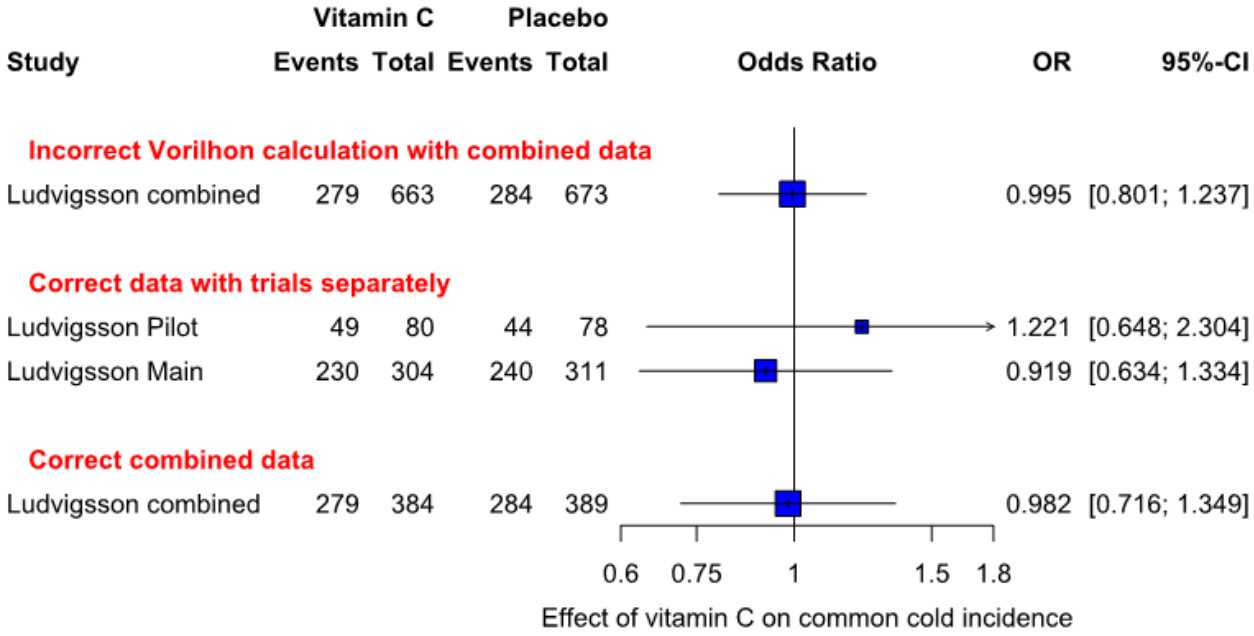


Calculation of the 95% CI for the Ludvigsson (1977) trials

In **Figure 10** below, Vorilhon's calculations are shown on the top and the correct published data and the corresponding correct OR and 95% are shown in the middle and at the bottom. This can be compared with Vorilhon's figure 2 which is shown on page 19 of this document. Although the point estimate in Vorilhon's calculation does not substantially differ from the correct point estimate (0.995 vs. 0.982), the confidence intervals are considerably different. The much wider CI in the correct analysis means that the trial has much less weight in the pooled analysis.

Furthermore, the two separate trials by Ludvigsson should be included in the meta-analysis as individual trials as shown in the middle of Figure 10. This would ensure appropriate weighting in the calculation. Combining the two distinct trials is not appropriate and in our own Cochrane analysis we included the two Ludvigsson trials as separate trials [2].

Figure 10: Comparison of correct presentation of the Ludvigsson's two RCTs in the middle with the incorrect combination by Vorilhon on the top and the correct combination on the bottom.



Part 3: Comments on Vorilhon's responses

Vorilhon's responses [5] to our first critique [3,4] are shown in yellow italics in this section. Our new comments are located after Vorilhon's comments. We have renumbered the references of Vorilhon's response [5] to generate a single numbering system for this document.

1. *Cohen et al.'s trial [13] appears very interesting in that it is the only one to focus on children of preschool age, a key period in the development of the immune system. Despite the significant preventive effects, it is currently difficult to conclude that echinacea [14] and propolis [15, 16] are effective in preventing upper respiratory tract infection (URTI) in children. However, even if the results of Cohen et al.'s trial are ignored in our analyses, the conclusions are identical, with odds ratio (OR) = 0.78 [0.54; 1.13] (p = NS) for the primary endpoint and standardized mean difference (SMD) = − 0.19 [− 0.31; − 0.07] (p = 0.002) (Supplementary Figs. 1 and 2).*

HH+EC: This is not a satisfactory response to our critique [3, 4 p. 2-3].

If Vorilhon is convinced that the administration of echinacea and propolis cannot confound trials on vitamin C, then it is his duty to show the evidence. Vorilhon does not refer to any studies to support the unambiguous inefficacy of echinacea and propolis. It is common wisdom that: "*absence of evidence is not evidence of absence*" [17,18].

Furthermore, the above statement
"*Despite the significant preventive effects, it is currently difficult to conclude that echinacea and propolis are effective in preventing upper respiratory tract infection (URTI) in children*"
is internally inconsistent, and that sentence does not support the belief in inefficacy.

Vorilhon's comment above that
"*However, even if the results of Cohen et al.'s trial are ignored in our analyses, the conclusions are identical*"
misses the point.

Inclusion of trials in a meta-analysis should not be justified by the findings of those trials. Inclusion of a particular trial should be justified by the stated inclusion criteria, and cannot be justified by the results of that trial. When Vorilhon [1] stated in the Methods section that "*RCTs comparing the use of vitamin C against placebo were selected*" the included trials should be restricted to trials in which the only difference between the treatment and control groups is the administration of vitamin C [4 p. 2-3].

[13] https://doi.org/10.1001/archpedi.158.3.217

[14] https://doi.org/10.1002/14651858.CD000530.pub3

[15] https://doi.org/10.1111/coa.12557

[16] https://doi.org/10.5455/jice.20160331064836

[17] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2550545

[18] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC351831

2. *In Coulehan et al.* [1974] *[8], random allocation into two groups is described as follows: "all children were assigned alternately, from an alphabetical listing by classroom, to one of two study groups". This sentence suggests an element of pseudo-randomization. In Table 6 (Risk of bias in the studies) [1], we assessed the risk of randomization in Coulehan et al. as 'unclear', because the word 'alternately' does not appear to be sufficiently clear. If we exclude Coulehan et al.'s trial, the conclusions are not modified for the primary endpoint: OR = 0.72 [0.50; 1.05] (p = 0.09) instead of 0.75 [0.54; 1.03] (p = 0.07).*

HH+EC: We don't feel our comments have been addressed.

Inclusion of trials should be consistent with the inclusion criteria. When the Methods section [1] states that "*RCTs comparing the use of vitamin C against placebo were selected*" the inclusion should be restricted to randomized trials.

Quasi-RCTs can be included, but if such trials are included in a meta-analysis, then the Methods section should state that quasi-RCTs were also included.

Inclusion of a trial should not be justified by the findings of the particular trial as Vorilhon suggests.

3. *Ludvigsson et al.'s article [6] contains reports of a pilot study and a princeps study, which we describe in Table 1. We did not exclude the pilot study of Ludvigsson et al. from the analysis for the two criteria 'incidence' and 'duration'. The data were pooled from the two studies (279/384 vs. 284/389, with 384 = 80 + 380 and 389 = 78 + 311).*

HH+EC: Ludvigsson (1977) [6] table V (**Figure 11** on the following page) describes the number of participants in the "Pilot study" and the "Main study" as follows:

> "Pilot study": 158 children: 80 vitamin C and 78 placebo
> "Main study": 615 children: 304 vitamin C and 311 placebo

Table 1 of Vorilhon's paper [1] has the title "Main characteristics of the included studies".
It describes the participants of the Ludvigsson (1977) trial as follows:

> 615 Swedish school children
> 304 vitamin C; 311 placebo

These figures in Vorilhon's table 1 are the "Main study" participants only (see our Figure 11). There are no figures in Vorilhon's table 1 describing the "Pilot study" participants, nor the total numbers mentioned in Vorilhon's above response "384" and "389".

It is possible that Vorilhon omitted a description of the "Pilot study" in their table 1, but might have included it in the meta-analyses.

That is clearly not the case in Vorilhon's meta-analysis on common cold **duration** in their Figure 3 [1] (see our Figure 11). Data for Ludvigsson's "Pilot study" is nowhere in Vorilhon's Figure 3.

Thus, Vorihon's statement "*We did not exclude the pilot study of Ludvigsson et al. from the analysis for ... 'duration' "* is not correct.

As to the Ludvigsson **incidence** data, Vorilhon's meta-analysis is based on incorrect data. There was no transparency in Vorilhon's meta-analysis on common cold incidence. However, we were able to reconstruct Vorilhon's calculations for Ludvigsson's incidence results. We found that sick people were double counted as sick and non-sick in the same meta-analysis (see page 22 of this document). The fact that the Pilot study was indeed included in the incidence analysis is irrelevant within an analysis containing so many errors, and errors even in the extraction of the Ludvigsson data. Finally, two separate trials should be analyzed separately and should not be combined in a meta-analysis (p. 26 of this document).

In the above response, Vorilhon calculated 384 = 80 + 380, but that is also not correct.

Figure 11: A copy of Ludvigsson's table V to describe the results for the Pilot and Main trials

Table V. *Occurrence of certain cold variables in control group and vitamin C group*

This above is from Ludvigsson (1977)

	Totally free from symptoms			Incidence (no. of cases/person)		Duration (no. of days/period)	
	N	%	*t*	M±S.D.	*t*	M±S.D.	*t*
Pilot study (30 mg, *N*=78; 1 000 mg, *N*=80)							
Cold symptoms from the nose							
30 mg	17	22	0.87	1.36±1.21	−1.59	7.61±8.07	1.82
1 000 mg	13	16		1.63±1.15		5.39±4.88	
Upper respiratory tract infection							
30 mg	34	44	0.68	0.71±0.72	−0.72	14.53±9.75	3.05**
1 000 mg	31	39		0.78±0.75		8.90±5.96	
Main study (10 mg, *N*=311; 1 000 mg, *N*=304)							
Cold symptoms from the nose							
10 mg	53	17	0.37	2.00±1.80	−1.41	5.67±7.89	−0.67
1 000 mg	49	16		2.16±1.63		6.04±5.47	
Upper respiratory tract infection							
10 mg	71	23	−0.35	1.28±1.03	−1.38	10.14±11.60	0.56
1 000 mg	74	24		1.39±1.11		9.54±8.65	

These "*Pilot study*" results are nowhere in Vorilhon's figure 3, see below.

Vorilhon's figure 3 [1] below gives the "*Main study*" results of Ludvigsson (1977), but there are no "*Pilot study*" results of Ludvigsson, though Vorilhon claimed so [5].

This below is from Vorilhon [1]

Study name	Mean ± SD	
	Treated	Control
Bancarali et al. (1984)	3.40 ± 2.77	4.50 ± 2.92
Cohen et al. (2004)	2.60 ± 4.20	6.20 ± 5.00
Constantini et al. (2010)	6.90 ± 5.40	8.90 ± 7.80
Coulehan et al. (1974, 1g)	4.95 ± 4.80	5.65 ± 6.40
Coulehan et al. (1974, 2g)	4.44 ± 4.80	6.29 ± 6.40
Ludvigsson et al. (1977)	6.04 ± 5.47	5.67 ± 7.89
Miller et al. (1977)	6.90 ± 10.00	7.50 ± 10.00
Ritzel et al. (1961)	1.80 ± 4.80	2.60 ± 6.40

Fig. 3 Forest plot: duration of upper respiratory tract infection

4. *We agree that the statistics section could have been more complete and we apologize for the lack of detail here. Where data were missing, we used imputation to estimate values, considering a conservative proposal, applying the higher standard deviation (SD) value reported in the metaanalysis for other studies that did not contain this information. This comment allows us to take better account of the data from Ritzel's trial [7].*

HH+EC: This response is ambiguous and does not reply to our comment.

Vorilhon's statement:

"Where data were missing, we used imputation to estimate values, considering a conservative proposal, applying the higher standard deviation (SD) value reported in the metaanalysis for other studies that did not contain this information."

suggests that Vorilhon may have looked for the highest SD value from the trials that did publish the SD value, and possibly used those values as the basis for the imputed SD values.

However, Vorilhon's figure 3 (see our Figure 2 on p. 4) shows that Miller (1977), Ludvigsson (1977), and Constantini (2010) had SD values higher than those imputed for the two Coulehan (1974) trials and the Ritzel (1961) trial.

Thus, Vorilhon's response above does not describe in any detail where the SD 4.8 days for the vitamin C group and the SD 6.4 days for the placebo groups originated for the Ritzel (1961) and the two Coulehen (1974) trials. They are not "the higher standard deviation (SD) value reported in the meta-analysis for other studies".

Although it is usually appropriate to be conservative in imputations when data are missing, the imputed figures should be consistent with the published data. That was not the case for the Ritzel (1961) trial in Vorilhon's meta-analysis as we pointed out in our critique [3,4] (see also p. 7 of this document). That was also not the case for the Miller (1977) trial, which likewise published data that can be used as a basis for the imputation of the SD values (see p. 10 of this document).

5. *The estimation of global effect was calculated according to the SMD estimates and simulations on SDs described for each trial. We agree (i) that a percentage scale approach could be seen to be more appropriate and (ii) that our proposition related to the SMD scale deserves to be discussed. According to the comment, the "26% effects" can be considered to be an overestimate.*

HH+EC: We don't feel our comments have been addressed.

In our criticism, we pointed out that Vorilhon wrote in their meta-analysis that *"the duration of URTI was decreased by 1.6 days (26%) in the vitamin C group"* [1, p 306], but the method of calculating these figures was not described at all. The statement *"… was found to decrease the duration of URTI by 1.6 days … p = 0.009"* was even published in the abstract and there should be a description of how the "1.6" was calculated.

The above response did not provide any further information on how Vorilhon actually calculated the 1.6 days and the 26% in their meta-analysis [1].

Vorilhon responds above that *"the '26% effects' can be considered to be an overestimate"*. However, there is no description in the above response as to what the basis is to claim that the 26% effect was an "overestimate".

Transparency is an essential part of science and the reader should be given information on how the calculations were actually carried out. That was missing from Vorilhon's paper [1] but it is missing also from Vorilhon's responses [5]. Furthermore, if a researcher changes his mind and considers that one figure was an overestimate, there should be transparency for the reasoning behind the change of mind. None is given in the Vorilhon response [5].

6. *Concerning the comment that the assessment of methodological quality could be considered inconsistent, it is important to reiterate that the trials are old and that there was no clear specification of whether they were carried out with 'intention to treat'; indeed, we think this to be unlikely. Even if the losses to follow-up were low, it is a limitation not to have information on whether this affects one group more than another.*

HH+EC: We don't feel our comments have been addressed.

We described in our critique [3, 4 p. 9] that in the Ludvigsson (1977) trial, 96% (615/642) of participants continued to the end of the "Main trial". In the Coulehan (1974) trial, *"Six hundred and forty-one of the 666 children (96 per cent) completed the entire 14-week study period"*.

Thus, the two trials had identical 96% dropout rates, but they were assessed differently by Vorilhon [1]. The above response does not give any justification or explanation as to why the two trials with identical drop-out rates were evaluated differently. The fact that both trials are from the 1970s does not change the requirement that trials with equal dropout rates should be classified consistently, unless there are explicit reasons for different conclusions, but none are given by Vorilhon.

7. *Finally, and concerning the 'manifestly incorrect data', we agree that, as suggested in these criticisms, this is more of a difference in the interpretation of results than an error, as mentioned previously. In Hemilä and Chalker's review [2], it is surprising that the main analysis was not clearly stratified by age, even if the authors discussed the effect of stress, whereas this was considered for the secondary criterion. Furthermore, we do not understand why fixed effect models were used, without taking into account the effects between and within studies (as opposed to random-effects models)*

HH+EC: Vorilhon writes in the response above:
"*In Hemilä and Chalker's review [2], it is surprising that the main analysis was not clearly stratified by age*".

In our 2013 analysis on vitamin C and common cold incidence there was uniformly no effect of vitamin C in adults and children in the general community, with no heterogeneity, after 5 small physical stress trials were moved to a separate subgroup [2]. When there is no heterogeneity, division by age does not give any useful additional information. Sometimes treatments differ by age or sex etc. and then it is informative to display the results by age or sex etc. However, if there is no indication that treatment differs by age or sex etc. then it is most informative to show the results for all such population groups together as we did [2].

It is not clear to us what Vorilhon means with the comment:
"*even if the authors discussed the effect of stress, whereas this was considered for the secondary criterion.*"

A meta-analysis of 3 RCTs in 1996 found that vitamin C prevented colds with participants under heavy acute physical stress, with the finding being statistically highly significant [19]. Therefore, separating general community trials and heavy acute physical stress trials was not a "secondary criterion" analysis in our vitamin C and common cold incidence meta-analysis in 2013 [2].

Our 2013 meta-analysis [2] included two later RCTs with participants under heavy acute physical stress [20,21], which were published after the 1996 meta-analysis [19]. Thus, those two later RCTs tested the hypothesis generated in the 1996 paper, that vitamin C may prevent colds in participants under heavy acute stress. The two later trials [20,21] supported the hypothesis and therefore the justification for subgroup analysis was even stronger than on the basis of the 3 RCTs in the 1996 paper [19]. Vorilhon does not describe what he means with the comment that heavy physical stress might be a "secondary criterion." When acute physical stress explains all heterogeneity [2], it is the most important criterion for subgroup analysis.

Vorilhon further writes: "*we do not understand why fixed effect models were used*"

First, Stephen Senn (2006) wrote:
"*it is always valuable to perform a fixed effects meta-analysis. This tests the null-hypothesis that treatments were identical in all trials. When and if this is rejected, then the alternative hypothesis that may be asserted is, 'there is at least one trial in which the treatments differed' "* [22, p 1427].
and Richard Peto (1987) wrote:

"*The random effects analysis says, 'look, we've got a lot of different trial results, here. What's the mean and what's the scatter of the different trial results?'. I think that this is actually wholly wrong as an approach to overviews and trials. I think it does answer a question. But it's a very abstruse and uninteresting question. It's trying to say 'what would happen if we chose another treatment at random from the universe of treatments that we could choose another population at random from the universe of populations'. I think this is not an important question. The question of interest, which I try to address by the standard p-value approach, is saying, 'Given the studies that people actually chose to do, have we observed more deaths in the treated groups that we would have expected just by the play of chance?'. I think that is the appropriate analysis, and that the*

random effects analysis is wrong; not statistically wrong, but commonsensically wrong. It's asking the wrong question" [23, p 242].

As an illustration of the different reasoning, when a random-effects meta-analyst is interested in the distribution of testicles in the community, he or she calculates that there is on average one testicle per person with a very wide confidence interval. In contrast, when a fixed-effect meta-analyst is puzzled with the great heterogeneity, he or she divides people by sex and reaches two separate fixed-effects estimates, both of which have very narrow confidence intervals. One testicle per person is a mathematically correct calculation, but it is an answer to a question that is biologically silly. It is asking the wrong question; not statistically wrong, but commonsensically wrong.

The random-effects meta-analysis approach often leads to combining of studies that are too different to be pooled. This is called the apples and oranges problem. Often it is much more informative to analyse potential explanations for the heterogeneity rather than pool inconsistent trials together [24]. For example, in our 2013 meta-analysis on vitamin C and the common cold incidence, there was significant heterogeneity over 29 trials (I^2 = 38% and P = 0.02; Analysis 1.1.1 in [2]). Dividing the trials into two subgroups, general community trials and heavy acute physical stress trials, explained all the heterogeneity. Comparison of the two subgroups is published at the bottom of the figure for that analysis: "Test for subgroup differences: Chi^2 = 22.74, df = 1 (P = 0.00), I^2 =96%" [2]. More accurately, the Chi^2 (1 df) = 22.74, corresponds to P = 0.000002. Thus, it is highly unlikely that the difference between the "heavy acute physical stress" and the "general community" trials is caused by spurious random fluctuation/variation.

Our approach is much more informative than combining all 29 trials and calculating the random-effects meta-analysis without trying to investigate reasons for the heterogeneity.

Second, in our meta-analysis on common cold incidence in the general population, there was no evidence of heterogeneity (I^2 = 0%, P = 0.77; Analysis 1.1.1 in [2]). In our meta-analysis on common cold incidence in the physically stressed participants, there was no evidence of heterogeneity (I^2 = 0%, P = 0.96; Analysis 1.1.2 in [2]). In our meta-analysis on common cold duration in adults, there was no evidence of heterogeneity (I^2 = 0%, P = 0.55; Analysis 2.1.1 in [2]). In our meta-analysis on common cold duration in children, there was also no meaningful indication of heterogeneity (I^2 = 27%, P = 0.17; Analysis 2.1.2 in [2]). When there is no heterogeneity, there is no difference in the results of fixed-effect and random-effects meta-analyses. Thus, even if we had used the random-effects approach within the above described subgroups, the findings would not have been any different.

In his response [5], Vorilhon gives the impression that our Cochrane review [2] is flawed because we used the fixed-effects meta-analysis approach. However, he does not describe what is wrong with the fixed-effect meta-analysis. Neither does he describe [5] how he thinks that the use of a random effects meta-analysis would have changed the conclusions of our 2013 review [2].

[19] https://doi.org/10.1055/s-2007-972864
https://helda.helsinki.fi/handle/10138/225881
https://www.mv.helsinki.fi/home/hemila/H/HH_1996_IJSM.pdf

[20] https://www.mv.helsinki.fi/home/hemila/CC/Moolla_1996.pdf

[21] https://journals.assaf.org.za/index.php/sajsm/article/view/914/599 see p. 23
https://www.mv.helsinki.fi/home/hemila/CC/Peters_1996_bm.pdf

[22] https://doi.org/10.1002/sim.2639 Senn p. 1427.

[23] https://doi.org/10.1002/sim.4780060307 Peto p. 242.

[24] https://doi.org/10.1001/jama.2019.5439
https://helda.helsinki.fi/handle/10138/312604